UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

February 16, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES #V-4

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich DK
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Michael Starsinic MS
Long Form and Variance Estimation Staff

Subject: Accuracy and Coverage Evaluation Survey: Specifications for Covariance Matrix Output Files from Variance Estimation for Census 2000

I. Introduction

The A.C.E. variance estimation operation will produce one official covariance matrix for national poststrata. This file will initially be used to compute small area variances which will be used to produce generalized variance parameters. These parameters will be published, and will be the way in which users approximate variances. All other internal users wanting variance estimates for verification, analysis or research will also use this matrix to get their measures of reliability.

This file will be a matrix of the covariances of the coverage correction factors (CCFs) for the 416 collapsed A.C.E. post-strata, which has been expanded to a 448 x 448 matrix. With this matrix, the only additional information needed to compute synthetic variances for any area or group are unadjusted census counts of the area or group broken down by the 448 collapsed A.C.E. post-strata.

$$\text{Var(A.C.E. Estimate)} = (C_1 \cdots C_{448})\, V \begin{pmatrix} C_1 \\ \vdots \\ C_{448} \end{pmatrix} = (C_1 \cdots C_{448}) \begin{pmatrix} V_{1,1} & \cdots & V_{1,448} \\ \vdots & \ddots & \vdots \\ V_{448,1} & \cdots & V_{448,448} \end{pmatrix} \begin{pmatrix} C_1 \\ \vdots \\ C_{448} \end{pmatrix}$$

$$= \sum_{i=1}^{448} \sum_{j=1}^{448} C_i C_j V_{ij}$$

where C_i = unadjusted census count for post-stratum i of 448
V_{ij} = covariance between post-strata i and j from CCF covariance matrix V
$\sum_i C_i$ = census count corresponding to A.C.E. estimate of interest

This file is for the United States (50 states and the District of Columbia) only. A file for Puerto Rico will be produced separately and discussed in a separate specification.

II. Specifications for the Matrix

• Platforms

This file will be stored on the VAX Alpha machine and, eventually, on the SRD Sun (UNIX).

• File Type

On both platforms, the file will be available as a SAS dataset (Version 6.12 format) and an ASCII text file which can be used in programs other than SAS.

• File Name

VCF448US: 'CF' for Coverage Correction Factor, and '448' for uncollapsed post-strata

• File Location

On the VAX Alpha:

CCFs, (448) Post-Strata
ESTIMATION$DISK:[VARIANCES]VCF448US.DAT
ESTIMATION$DISK:[VARIANCES.SDS]VCF448US.SASEB$DATA

• Layout

All the files will be in a standard linear format. The first two columns will be the post-stratum identifiers (numbered 1-448) for post-strata i and j, respectively. The third column will be the covariance of post-stratum i and j.

CCF Variance ASCII File Formats			
Variable	Description	Columns	Format
i	Post-Stratum i (of 448)	1-3	I3
j	Post-Stratum j (of 448)	5-7	I3
Cov	Covariance of CCF_i and CCF_j	9-25	F17.14

III. User Verification

To reduce the possibility that the matrix will be used incorrectly, we will include separate files with precalculated "benchmark" variances, together with the post-strata counts used to compute them. The user will be requested to run their variance calculation program to reproduce these benchmark variances. If the users' computation fails to match the benchmark values, then there was an error somewhere in their code. There is no way to require users to run this benchmarking, but we will strongly encourage them to do so to decrease the chances of miscoding their variance estimation program and producing erroneous results.

• Verification Files

For the VCF448US file, there will be four associated benchmark files:

.test1	Contains the US total in-scope population (excluding GQs & Remote Alaska) by the 448 collapsed post-strata.
.answer1	US national variance, as calculated from the covariance matrix and the .test1 file.
.test2	Population file to calculate a variance based on a random subset of the US.
.answer2	Resulting variance from covariance matrix and .test2 file.

• File Location

On the VAX Alpha:

ESTIMATION$DISK:[VARIANCES.BENCHMARKS]VCF448US.TEST1
ESTIMATION$DISK:[VARIANCES.BENCHMARKS]VCF448US.TEST2
ESTIMATION$DISK:[VARIANCES.BENCHMARKS]VCF448US.ANSWER1
ESTIMATION$DISK:[VARIANCES.BENCHMARKS]VCF448US.ANSWER2

• Layout

Both of the .test* files will be in a standard linear format, giving each of the 448 population counts or indicators on a separate line. The .answer* files contain a single line with the resulting variance.

VCF448US.test* File Formats			
Variable	Description	Columns	Format
Pop	Population of PS i	1-7	I7

VCF448US.answer* File Formats			
Variable	Description	Columns	Format
Var	Variance of VCF448US.test*	1-16	F16.3